

Quick guide to GoDig

Qing Yu,¹ Xinyue Liu,¹ Mark P Keller,² Jose Navarrete-Perea,¹ Tian Zhang,¹ Sipei Fu,¹ Laura P Vaites,¹ Steven R Shuken,¹ Ernst Schmid,³ Gregory R Keele,⁴ Jiaming Li,¹ Edward L Huttlin,¹ Gary A Churchill,⁴ Devin K Schweppe,⁵ Alan D Attie,² Joao A Paulo,¹ Steven P Gygi^{1,*}

¹Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA.

²Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706, USA

³Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA.

⁴The Jackson Laboratory, Bar Harbor, Maine 04609, USA

⁵Department of Genome Sciences, University of Washington, Seattle, WA 98105, USA

Download Tomahto

- Prerequisite: Thermo API license (<https://github.com/thermofisherlms/iapi>)
- Download website: <https://gygi.med.harvard.edu/software> (also smarttmt.org).
- Orbitrap Tune software: 3.5 or later version

Smart-TMT

Real-Time-Search

The Real-Time Search** (RTS-MS³) provides real-time (<5 ms / spectrum) spectral identification and triggers SPS-MS³ scans that utilize assigned and pure fragment ions for accurate quantitation. Time consuming SPS-MS³ spectra are only acquired after confident peptide identification, greatly increasing the number of peptides interrogated and reducing the effects of isobaric interference. **Try it!**

Tomahto

The Tomahto software provides real-time instrument control and decision making. Tomahto enables simplified implementation of TOMAHAQ targeted assay. It provides an array of functionalities including MS1 peak detection, MS2 real-time peak matching (RTPM), MS2 fragmentation pattern match, SPS ion purity filter, MS3 automatic gain control (AGC), MS3 quant scan insertion, and target peptide close-out. In addition to controlling data acquisition, it also allows real-time data visualization and post-acquisition analysis. **Try it!**

GoDig

The GoDig software is a real-time analytics platform that enables next-generation TMT-based multiplexed targeted proteomics. It obviates the need for making internal peptide standards and tedious method curation. As a result, it essentially makes targeting nearly any previously detected peptides possible. One only needs a list of peptides and GoDig will perform real-time elution calibration and spectral matching to identify the targets, and prompt SPS-MS3 scans to quantify them down to attomole level. **Try it!**

General considerations for trigger peptide selection

1. Should be unique to the protein of interest
2. Should avoid missed cleavages to minimize artifact caused by sample preparation
3. Should not contain methionine to avoid artifact caused by oxidation
4. If a peptide contains cysteines, an extra step needs to be taken to reduce/alkylate the free cysteines using the same protocol as the endogenous protein sample preparation
5. Preferred peptide length is 8~20 amino acids to ensure proper hydrophobicity
6. We prefer to have 2 peptides per target protein as a fail-safe measure

How to acquire data using GoDig.exe

- 1 Connect instrument
- 2 Select target peptide list (.csv)

Example of input .csv file

GeneSymbol	Peptide
Abhd1	YTSVAFGYK
Abhd1	VLDVDFAIK
Abhd10	SELPNLAYK
Abhd10	IPYSFIK
Abhd11	LNLDTLAQHLDK
Abhd11	TNFNSLAK
Abhd11	VLTVDAR
Abhd12	VPYFIDLK

- 3 Select peptide spectral library (.xml)
- 4 Select peptide elution library
(e.g., demo_elution_library_4_cell_line_tmtpro.bin)
- 5 Select proteome database
(.idx; e.g., demo_uniprot_HUMAN_TrypsinKR_TMTPro.fasta.idx; see next slide about how to build one)

- 6 Select elution calibration interval and desired bin width
- 7 Select peptide modification and data acquisition parameters (see next slide)
- 8 Load parameter. This may take a couple minutes.
- 9 Start acquisition

The screenshot shows the GoDig software interface with the following components and numbered callouts:

- 1**: Connect button (top left)
- 2**: Load Parameter button (top left)
- 3**: Start Acquisition button (top left)
- 4**: Scans Arrived field (top right)
- 5**: MS2 / MS3 fields (top right)
- 6**: Closed out field (top right)
- 7**: Peptide tab (middle left)
- 8**: MS Method tab (middle left)
- 9**: Real Time View tab (middle left)
- 1**: Instruction list (middle left)
- 2**: Target Peptides section (middle left)
- 3**: Targets (.csv - Columns = "Peptide.z.Gene Symbol") field (middle left)
- 4**: Spectral Library (.xml) field (middle left)
- 5**: Elution Calibration section (middle left)
- 6**: Elution Library (.bin) field (middle left)
- 7**: Comet Database (.idx) field (middle left)
- 8**: Bin Setting: Bin Width (±min) 0.5, # Bins (±) 3 (middle left)
- 9**: Calibration Setting: Cal. Interval (s) 15, Top N Peak 6, Bin Offset 1 (middle left)
- 1**: Modifications table (right)
- 2**: Target Peptide table (bottom right)

Modifications Table:

Target	Name	Mono Mass	Type	Sites
☐	TMT0	224.15248	Static	K,NPep
☒	TMT11	229.16293	Static	K,NPep
☐	TMTsh	235.17677	Static	K,NPep
☐	TMTpro0	295.1896	Static	K,Np
☐	TMTpro18	304.2071	Static	K,Np
☐	CTerm-K602	8.0142	Static	PepC-K
☐	CTerm-R604	10.00827	Static	PepC-R
☒	CAM	57.02146	Static	C
☐	NEM	125.04767	Static	C
☒	OX	15.99491	Dynamic	M
☐	Phos	79.96633	Dynamic	S,T,Y

Target Peptide Table:

Peptide	Z	Gene	Target M/Z	CV
*				

How to acquire data using GoDig.exe

- 1 Select scheduling method
(default to elution order)
- 2 Set dynamic exclusion
(how frequent a target should be monitored)
- 3 Set trigger method
(default to PRM)
- 4 Set MS parameters
- 5 Choose directory to write log file
- 6 Choose close-out option if needed

The screenshot shows the GoDig.exe configuration window with the following sections and settings:

- Schedule Method** (1): Method Length (min) is 120. ☐ None, ☐ RT Window (±min), ☒ Elution Order.
- Dynamic Exclusion** (2): Tol. (±ppm) is 7, Duration (s) is 5.
- Trigger Method** (3): ☐ MS1, ☒ PRM.
- MS2 Parameter** (4): Min. Intensity (MS1 trigger only) is 50000, RF Lens (%) is 30, Isolation Width is 0.5, Activation Type is CID, MSA is ☐, Neutral Loss is 97.9763. Trigger table:

	Trigger	Target
NCE	35	35.1
Max IT	120	900
AGC	10000	100000
Analyzer	IonTrap	Orbitrap
IT Scan Rate	Normal	
OT Reso.		15000

Filters: Trigger Frag. Tol. (±) is 0.15 (ppm), Target Frag. Tol. (±) is 15 (ppm), Match Top N Frags. is 10, Min. Matched Peaks (OT) is 3, Min. Cosine Score is 0.90.
- SPS-MS3 Property** (4): Prec Exclusion Window is 50 - 5, SPS Ion Range (Th) is 400 - 2000, SPS Above Precursor MZ is ☐, Num. SPS Ion is 4, SPS Ion Cutoff (% of BP) is 2, MS2 Iso. Width is 0.8, MS3 HCD CE is 65, MS3 OT Res. is 50000, Insert prescan is ☒, MS3 AGC is 250000, MS3 MaxIT is 2000.
- Log File Option** (5): Log Export Folder is empty, Start Writing Log button.
- Close-out** (6): Stop after sum S/N AND # MS3s reached: Sum SN is 1000, AND # of Scans is 1. Delay close-out until failed monitor/ID: ☐ Stop after N failed post-quant monitor scans (3), ☐ Stop after first failed post-quant ID MS2. Multi-run options: Unclose targets after injecting vial is 1-V4, Wait until end of run to close out ☐, Unclose targets at end of each run ☐. Start Close-out button.

Server Connected: False | System Mode: | System State: | Instrument Name: | Instrument Connected: False

How to index a proteome .fasta file to be used for elution calibration

- 1 Select proteome fasta file
(can be obtained from Uniprot <https://www.uniprot.org/>)
- 2 Select TMT labeling
- 3 Select search parameters. GoDig uses Comet as its real-time search engine. Please refer to Comet website for all details regarding the parameters (<https://comet-ms.sourceforge.net/parameters/>)
- 4 Index database. A .idx file will be generated in the same folder as the .fasta

Peptide MS Method Real Time View Data Analysis Elution and Spectra Library Database

Database Setup

Searching with Orbiter requires an indexed database. Indexing a database only needs to be done once per parameter set (e.g. yeast, with TMT and phos).

Decoy proteins (label: "###") required for SMART filtering.

(1) Choose FASTA File

(2) Append Text to New Database File:

☒ Delete Temporary Fasta

(3) Assign Database Parameters

☐ Label Free ☐ Amino-TMT6/10/11 ☐ Amino-TMT11 PHOS
☐ Amino-TMT0 ☐ Iodo-TMT0 ☐ Amino-TMTPro
☐ Amino-TMT2 ☐ Iodo-TMT6

Parameter	Value	Example
database_name	EXAMPLE.fasta	'###' specifying reverses
search_enzyme_number	1	Default to trypsin (1)
peptide_length_range	7 63	Exclude small peptides fo...
max_duplicate_proteins	20	-1 reports all duplicates
variable_mod01	15.9949146221 M 0 3 -1 ...	MetOx: 15.9949 M 0 3 -1 ...
variable_mod02	0.0 X 0 3 -1 0 0 0.0	0.0 X 0 3 -1 0 0 0.0
variable_mod03	0.0 X 0 3 -1 0 0 0.0	0.0 X 0 3 -1 0 0 0.0
fragment_bin_tol	1.0005	IT: 1.0005; OT: 0.02
fragment_bin_offset	0.4	IT: 0.4; OT: 0.0
theoretical_fragment_ions	1	IT: 1; OT: 0

(4) Index FASTA Database

Data analysis

- 1 Select peptide list (.csv) and raw file.
Note: you have to also select same modification, and fragment match tolerance in the method editor as the rawfile was acquired with.
- 2 Select raw file (.raw)
- 3 Select spectra library (.xml)
- 4 Process
- 5 Export to .csv

GoDig

Stats
Scans Arrived: MS2: /
Closed out: MS3: /

9:38:57 AM Start Extracting XICs...
9:39:11 AM Closing Raw File
9:39:11 AM Done!

Peptide MS Method Real Time View Data Analysis Elution and Spectra Library Database

Target List: 1 Target list (.csv) Browse Analyze Data 4
Raw File: 2 Raw file (.raw) Browse
Spec. Lib.: 3 Spectral library (.xml) Browse Bulk Process

Search Target:

Select	Peptide	z	GeneS	Target
☐	HFPNIDR	3	DYN...	401.5582
☐	DTIEEHR	3	WA...	401.8811
☐	LHTFESHK	4	RBB...	402.4854
☐	DLEGLSQR	3	MYH9	407.8968
☐	ILMEHIHK	4	RPL...	408.0005
☐	GLVLDHGAR	3	CCT...	414.581
☐	ELSDIAHR	3	ALD...	415.5688
☐	HIAEDSDRK	4	TPM2	420.4896
☐	HGLYLPTK	3	FASN	420.9176
☐	MEIYRPHK	4	HDA...	421.248
☐	TFHHVYSGK	4	DDX...	421.742
☐	ERHPGSFDVVHVK	5	RPS...	423.8449
☐	ASLPGVK	3	EEF...	427.2791

5 Export Data

☐ Display Only Targets with Data 930/938 genes have data
☐ Display Selected Only 2571/2600 peptides have data

Server Connected: False System Mode: System State: Instrument Name: Instrument Connected: False

Graph

Precursor XIC

Intensity

Retention Time (min)

Library Spectrum

Relative A...

m/z

Trigger MS2

IT: 50 ms

Target MS2

IT: 100 ms; Cosine: 0.99

Reporter Ion SN

Sum S/N: 5112

TMT Channel

MS/MS Events

	Include	MS2 RT	Precursor Int (Log10)	MS2 Scan#	MS2 IT	MS2 RT	MS2 Scan#	MS2 IT	Trig Peak	Targ Peak	Cosil	Corr Frag	Com Inter Frag	BP Pres	Frac of Frag	# SPS	Qu
▶	☒	16.65	0.0E+0	8472	100	16.66	8479	150	14	13	0.99	9	100	True	10...	6	24
	☒	16.74	0.0E+0	8523	100	16.78	8537	150	14	14	0.99	9	97...	True	10...	5	32
	☒	16.94	0.0E+0	8622	100	16.96	8633	150	13	10	1.00	9	97...	True	10...	5	8.6

How to build custom library

- 1 Select peptide modification
- 2 Select which library to build (e.g., spectral and elution)
- 3 Select the .csv file that contains all identification information to be used for the library (see next slide for details)
- 4 Additional filters to remove low score peptides or peptide with large elution window. Fragment mass tol need to be set to build spectral library.
- 5 Raw files that are used to generate the .csv search result
- 6 Start to build (this may take a while, depending on the size of the dataset)

Modifications

Target	Name	Mono Mass	Type	Sites
☐	TMT0	224.15248	Static	K,NPep
☒	TMT11	229.16293	Static	K,NPep
☐	TMTsh	235.17677	Static	K,NPep
☐	TMTpro0	295.1896	Static	K,Npеп
☐	TMTpro18	304.2071	Static	K,Npеп
☐	CTerm-K602	8.0142	Static	PepC-K
☐	CTerm-R604	10.00827	Static	PepC-R
☒	CAM	57.02146	Static	C
☒	NEM	125.04767	Static	C
☒	OX	15.99491	Dynamic	M
☐	Phos	79.96633	Dynamic	S,T,Y

Target Peptide

Peptide	Z	Gene	Target M/Z	CV
*				

Peptide **MS Method** **Real Time View** **Data Analysis** **Elution and Spectra Library** **Database**

Build Elution/Spectra Library

1. Choose to build elution library and/or spectra library.
2. Load .csv output from Core.
3. Select PSM filter criteria.
4. Select all rawfiles for the search results.
5. Click "Process". Xml file(s) will be stored in the same folder as the Core output.

Build **Options**

☐ Elution Library ☐ Spectra Library

Min. Score Max. Peak Width (min) ☒ Use Best PSM

Frag. Mass Tol. (±) ☒ DA ☐ PPM

Background Search Result

(.csv File: ScanF, Time, z, Parent Scan, SrchID, SrchName, Peak Width, Gene Symbol, Reference)

Raw Files

How to build custom library--.csv file

Example csv input:

Peptide	z	Theo m/z	Parent Scan	Reference	ScanF	Expect	Time	SrchID	SrchName	Peak Width	Gene Symbol	PepID
R.QQEGESR.L	2	569.2946	786	sp P00338-3 LDHA_HUMAN	795	6.592	5.3	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.3003	LDHA	43
R.SQEAGGR.V	2	504.7733	786	sp P48509 CD151_HUMAN	800	5.252	5.31	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.3003	CD151	48
K.AGQGSGSR.K	2	512.2787	794	sp A4FU01 MTMRB_HUMAN	808	6.686	5.33	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.3093	MTMR11	55
K.NQDDADR.A	2	569.2764	794	sp O75937 DNJC8_HUMAN	813	7.759	5.33	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.287	DNAJC8	60
K.REQAEER.Y	3	450.9026	817	sp Q9UII2 ATIF1_HUMAN	826	5.471	5.36	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.2783	ATP5IF1	71
K.RAQEEAER.L	3	431.5674	817	tr E7EQR4 E7EQR4_HUMAN	827	6.342	5.36	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.3003	EZR	72
K.NEEDEGHSNSSPR.H	3	587.9368	817	sp Q14103 HNRPD_HUMAN	830	11.026	5.36	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.2565	HNRNPD	75
R.CNDQDTR.T	2	606.7835	817	sp P02751-15 FINC_HUMAN	833	8.64	5.37	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.2565	FN1	78
K.NASCGTR.S	2	535.2726	825	sp P60468 SC61B_HUMAN	838	4.764	5.37	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.3093	SEC61B	82

Peptide: peptide sequence with or without two flanking residues (e.g., R.QQEGESR.L or QQEGESR)

z: peptide charge state

Theo m/z: theoretical peptide m/z

Parent Scan: master scan the PSM is based on

Reference: Uniprot reference

ScanF: Scan number of the PSM

Expect: Expect score from Comet search or XCorr from Sequest search (Can be other score too but need to use the same column header)

Time: Scan time

SrchName: raw file name without the file suffix (e.g., ec04297_qy_GoDig_4cell_CIDhrMS2 from ec04297_qy_GoDig_4cell_CIDhrMS2.raw)

Peak Width: peptide peak width

Gene Symbol: gene symbol

PepID: each PSM from one raw file has a unique peptide ID. If not provided by a search engine, **ScanF** can be used instead since it is unique in each raw file.

More questions?

Please email Qing Yu (qing_yu2@hms.harvard.edu) or Steven Gygi (steven_gygi@hms.harvard.edu)